Evaluation of Federated Searching Options for the School Library

Sarah E. Abercrombie is a Librarian at the Greenwich (Conn.) Country Day School and a student at Southern Connecticut State University, New Haven. She is finishing her MLS and certification as a media specialist.

Three hosted federated search tools, Follett One Search, Gale PowerSearch Plus, and WebFeat Express, were configured and implemented in a school library. Databases from five vendors and the OPAC were systematically searched. Federated search results were compared with each other and to the results of the same searches in the database's native interface to disclose differences in handling query syntax, searching, retrieval, browsing results, etc. Each product was easily configured, but none were capable of searching every database desired. Simpler Boolean queries are the most successful queries because of the underlying structure and differences of the databases, and the capabilities of certain products. Federated search products succeed in simplifying access to multiple database resources at school, but searching remains different from the familiar Web search engines in many ways. To become more Google-like, federated searching must be done against indexes built in advance instead of the current real-time searching method.

Libraries are faced with formidable competition from Google and other Web search engines to be their patron's information resource of choice. Students could be using the accurate, authoritative, and age-appropriate print and electronic sources provided by school and public libraries for their school research, yet, according to a Pew study, the first place most students turn to for their information is the Internet via Google, or other Web search engines (Levin and Arafah 2002), and they have tremendous faith in their favorite search engine (Tenopir, Hitchcock, and Pillow, 2003). Many studies have shown that school students favor the convenience and speed of access of using electronic information resources for school-related work (Ibid.). School libraries like Greenwich (Conn.) Country Day School (GCDS) have responded by buying less print and investing more of their budgets in age appropriate subscription database and electronic reference materials. These databases provide advantages over the Internet by concentrating grade level content and providing citation services. Using trusted subscription database resources instead of Internet sources might help address the weaknesses children have using the Internet as a source for schoolwork summarized in Large's (2004) meta-analysis of children's information seeking on the Internet. Large reported that children were able to judge relevance on the basis of topicality, novelty, and interest, but did not question accuracy, authoritativeness, or truthfulness.

Unfortunately, providing access to quality electronic information is not enough to convince school students to use library resources over the Internet. A 2005 OCLC study of college students' perceptions of libraries also included a section studying future college students' perceptions (fourteen to seventeen year olds). This section showed that libraries lack relevance in the lives of the high-school age respondents, and that, for this age group, library resources and services are not clearly differentiated from other information sources. (Da Rosa et al. 2006). The fact that students often don't differentiate quality library resources versus questionable Internet resources is one problem, but another very serious problem is that students quickly give up trying because searching library resources is different and more difficult than searching the Internet.

Students bring their Web searching habits to their use of all electronic resources (Tenopir et al. 2003). They expect to be able to simply enter some words into a search engine and get their answers—not an unreasonable expectation in the era of Web search engines. "For the Google generation, the quality of results is not as important as the process, which must be simple, quick and efficient" (Labelle 2007, 241). Compared with the speed and simplicity of a Google search, which is many students' model of what an information search should be, library electronic resources are slower, more difficult, and inefficient. In most cases, a patron must first choose which databases would be best for their needs—often from alphabetical lists of databases. Once selected, most databases will be queried individually from their own unique search interface. There is an authentication step if the user is accessing from outside school, requiring username and password combinations. Federated search tools strive to provide users with a single point of access to retrieve relevant, quality content from multiple database resources with one simplified search. Federated searching is the process whereby a search engine connects to multiple information resources simultaneously, broadcasts a query to each database in real time, then retrieves the results and displays them to the user. With substantial investments in subscription databases, schools are looking at federated searching tools to simplify and thus increase database usage by providing a single point of access and at the same time decrease student reliance on Web search engines.

Librarians have debated the merits and drawbacks of federated search tools since they first appeared in the academic library. They are imperfect search tools that have been accused of "dumbing down searching" (Fahey 2007, 62) or being a step backward, a way of avoiding the learning process (Frost 2004). There is a perceived negative effect on student's information literacy. The sophisticated features available when searching a database directly from their own (native) interfaces such as thesauri, subject headings, and limiting tools are lost when searching from a federated searching tool, but most library users currently do not have, or choose to not employ, the skills to use these kinds of enhanced searching tools (Fahey 2007). Students prefer ease-of-use and speed over relevance and depth. They have little patience, and are more likely to opt for single search tools than replicating searches in multiple search tools. They prefer keyword searches and usually ignore interface enhancements and alternative search options, favoring simplicity over sophistication. They seem satisfied with the first reasonable results, and rarely seek comprehensiveness (Labelle 2007). Proponents of federated searching say that libraries

should make finding information easier, and until libraries are able to connect with and teach all users the complexities of the search process so they understand the depth and breadth of the resources available, federated searching should be offered. Although users will miss some quality results with a federated search, they will find some that would not otherwise have been located. They argue that a common interface that mirrors a web search engine could wean students from their dependence on the Internet (Frost 2007, Labelle 2007).

The terms "federated searching," "metasearching," "integrated searching," "cross-database searching," "parallel searching," and "broadcast searching" are often used interchangeably (National Information Standards Organization [NISO] 2006) to describe the process whereby a search engine connects to multiple licensed or free resources, broadcasts a query to each database simultaneously in real time, then retrieves the results and displays them to the user. That definition is the basis for this paper, though distinctions between terms are sometimes made. Solomon (2004) suggests that federated searching is distinguished from integrated searching by a system's ability to post-process the retrieved results to somehow rank, cluster, or categorize them. Using this definition, only Gale's Power Search Plus and Webfeat Express would be considered "federated" searches in this study. Sadeh (2007) and Freund, Nemmers, and Ochoa (2007) make the debatable distinction between federated "just-in-case" searching and "just-in-time" metasearching. They submit that, in federated searching, the system harvests and processes the metadata from a repository in advance, prior to a user's search; a process analogous to Google's use of robotic crawlers to harvest information from the Web to build indexes, which are then searched against users queries. They consider metasearching to be real-time. This distinction is not popular in the literature because presently none of the leading vendors of federated search products index the databases they search. They use software connectors to perform searches in real time against an actual database. Software connectors are computer coding that transmits a user's query (plus any authentication information from the federated search engine's query box to a query in each selected database) then collects the results and transfers them back to the federated interface. Sadeh may be making the distinction on the basis of what she and others (Breeding 2005) see as the future of federated searching.

Breeding states that the current strategy of metasearch that relies on live connections can never stand up next to search systems based on indexes created in advance, and that "an entirely new approach is needed" to provide a simpler, faster search and retrieve protocol (2005, 28). He advocates a centralized search model that harvests metadata into indexes that can provide instantaneous results to user queries; the model used by Web search engines like Google. The centralized model is now economically possible because very large-scale data storage and clusters of computers to deliver almost limitless computer power are more affordable than ever. But he stated that having publishers expose their entire collections for metadata harvesting and document indexing hadn't yet been practical from a technical or business perspective. This is changing gradually as publishers increasingly open their content to Web crawlers as evidenced by the growth of Google Scholar, which indexes articles from many proprietary publisher databases.

A review of the literature from 2002 to the present contained in Emerald Library Suite, Library Literature, Library and Information Science Abstracts, Library and Information Technology Abstract, Academic Search Premier, and ERIC databases was conducted using keywords such as "federated," "metasearch*," "meta-search* database," "cross-database searching," as well as the names of schools, libraries, vendors, and product names. In addition, studies and meta-analyses of student use of electronic information sources (Tenopir et al. 2003), the Internet (Levin and Arafah 2002, Large 2004), and libraries (Da Rosa et al. 2006) were reviewed to support the study.

Federated searching has been studied almost exclusively in the context of academic libraries. The literature contains many studies of the evaluation and implementation of various federated search platforms such as MetaLib from ExLibris, ENCompass from Endeavor, and WebFeat in the academic library. This finding is confirmed by Freund, Nemmers, and Ochoa's (2007) annotated bibliography of forty-four journal articles on metasearching. Almost all of the articles focused on some aspect of academic library implementation, including twenty-two articles focused on the evaluation or implementation, including usability testing, of federated search vendor products, other articles comparing federated search with Google or Google Scholar, and studies of the effect of federated searching on information literacy. Only three brief articles were found—Minkel (2002), Curtis and Dorner (2005), and Young (2005)—that discuss cross-database searching in the school library context. These articles address cross-database searching capabilities within vendor families such as Gale, EBSCO, and ProQuest, but not the capability to search a broad range of databases from a variety of vendors and the OPAC from one federated search.

The Federated Search Process

Although federated search products differ in many ways, the selected federated products are all hosted products, meaning that all of the computing and processing of results takes place, or is *hosted*, on the federated search vendor's computers, so no equipment or programming is required for the library. All of the products offer similar functionality to users including resource discovery, querying multiple databases, and retrieving and displaying results. They are each configurable by the librarian, but differ in the features and amount of customizing that is available.

Resource discovery. First, federated search engines enable users to select the resources they want to search. Being selective about which sources to search increases the relevance of the retrieved articles and increases the speed of the search, not to mention decreasing the burden on the database provider's infrastructure. WebFeat Express, a more expensive and sophisticated federated search product, offers the capability of grouping databases into preselected subject area sets to help novice searchers discover what databases might be best for their area of research. All of the products allow users to select multiple resources to search and have a librarian-configured default set of databases.

Querying multiple databases. The user enters query terms into either a single search box or an advanced interface with pull-down Boolean operators and in some cases pull-down field name boxes. The federated search engine transfers the query to the selected resources in real time with the help of unique software connectors specific to each database selected. The software connectors take care of any authentication required, and in some cases translation of the query into a form that complies with features of the search engines of each target resource.

Waiting for Results. Executing a search in a database takes time. Retrieval times suffer because of bandwidth bottlenecks from the network leading to each resource and the performance of the server processing the results (Sadeh 2007). Typically, databases will only return results in batches of twenty or thirty at a time and will cut off a search when a certain number of hits are retrieved. The speed of the slowest database being searched determines the fastest speed of the federated search. Because users expect almost instantaneous results, Federated search systems often display the first batch of results almost immediately. This is helpful since users can often determine if their query was appropriate based upon initial result set.

Browsing and Displaying Results. Federated search systems vary in the way retrieved results are handled. In the most basic products, results are grouped by source, returned in the order they are received, and displayed with a brief description of the item and a link to either more metadata within the federated system or directly to the item in the target resource. More sophisticated systems will process the retrieved results from the various search engines in an attempt to rank, cluster or categorize results. Systems also vary in browsing features such as sorting by source or relevance or displaying results visually or in list form.

Obtaining Services. There are also other valued services that federated systems may offer, including customizable personal workspaces where a user can save result sets, search histories, set preferences, specify alerts, and so on. In the academic library, where abstracting indexes are often searched, federated search systems can further the objective of finding full-text articles by obtaining services from article-linking services such as ExLibris SFX. This functionality is not relevant to a school-oriented federated service because school libraries tend to subscribe only to databases that are either entirely or mostly full-text.

A variety of vendors have offered customized and complex federated search tools for academic libraries beginning in 1998 (WebFeat n.d.), but these customized, expensive federated products were not feasible for a school library due to the technical complexity of installation and maintenance and the cost. This study compares the implementation, searching capabilities, retrieval, and functionality differences of a new generation of hosted metasearch tools: Follett One Search, Gale PowerSearch Plus, and WebFeat Express, implemented in a school library environment. A hosted federated search is conducted on the vendor's computing hardware, reducing the cost and complexity to a workable level for a school library.

Method

Test Environment

Students and faculty at GCDS have access to more than fifty electronic information databases through either direct subscription, or the iConn consortium. These databases include resources suitable for early readers through professional educators. The databases are provided by ten different vendors, and, with some exceptions within vendor families, each database has its own unique proprietary interface. Students and faculty have ubiquitous access to computers; students in grade 7–9 have their own laptop computers, and younger students have ample computer access in classrooms and labs. Librarians actively promote the use of electronic resources by students and faculty.

Federated Search Product Selection

The federated product selection was narrowed to products that were Web-hosted, were ready to use without special vendor customization, were configurable by a typical librarian with some experience managing electronic resources, and were capable of a comprehensive search of both electronic databases and print materials via the OPAC. Many library-management system companies offer some cross-database searching capability with their OPAC products at an additional cost.

GCDS uses Follett Destiny as the integrated library management system, and they also purchased their cross-searching One Search tool module. As a Gale customer familiar with the PowerSearch product that cross-searches some, but not all, of the databases within its own product line, GCDS purchased their new federated offering PowerSearch Plus when it went online in January 2008. The third federated search tool is WebFeat Express, which also met the criteria, and was secured on a thirty-day trial.

Resource Selection and Configuration

In order to be able to compare the data from the systematic search using the three search tools, all three products had to be configured to search the same selection of electronic resources. Each federated

search product has a list of supported databases and resources to choose from. One selects a free resource (such as Internet Public Library) or subscription database vendor family (such as ProQuest, Ebsco or Gale) from a list, then chooses the specific database from a list of supported databases. Authentication information—I/P and/or password—is entered, and in each case, a preference for searching the database by default is set. One of GCDS's most highly used database products from each of the major vendors as well as commonly used electronic resources such as encyclopedias and the OPAC were selected. The Gale databases had to be chosen from a list of databases not searched in their regular PowerSearch platform, and it took a few revisions to arrive at a list that was workable for all three products. The list included Gale's History Resource Center (U.S. and World), Ebsco's History Reference Center, ProQuest's Historical NY Times (HNYT), the Encyclopedia Britannica, the World Book Encyclopedia, and the OPAC (Follett Destiny).

Systematic Search Process and Data Collection

The search process was designed to reveal how the three federated search products interpret queries versus the native interface provided by the database vendor, and if there were any differences among them that would affect retrieval of results. Searches were conducted in groups of two to four searches using starting queries similar to those that students have used in conducting research in the media center. Each of the starting queries were modified slightly to include or exclude a search term or syntax element and re-executed to determine the effects of the change on retrievals. The effect of Google-type syntax and familiar search engine operators were studied, as well as a selection of typical database queries such as subject heading and author searches.

Treatment of Results Returned

Search-results data from identical queries were collected from each federated product and each native interface and tabulated. Federated results were grouped by source because although PowerSearch Plus can return results by source, by cluster (default for outline view), or by relevancy ranking (default list view), and WebFeat can return results by source, title, author, date, cluster, or relevancy, Follett One Search can only return results grouped by source. The number and to some extent the order of results returned were compared and analyzed for differences. Although a comparison of relevancy ranking algorithms is out of the scope of this paper, article ranking was examined qualitatively during the searches to get a feel for the utility of relevancy ranking in these federated search tools.

Results

Federated searching delivers on its promise of easily searching multiple resources with a single query, but fails in some ways to make database searching more Google-like. Many of the differences between searching the Web and searching subscription databases that users find frustrating or confusing remain because of the structure and syntax requirements of the underlying databases and the time delays of real-time searching. Different databases treat multiple word queries differently. The Encyclopedias treat multiple words as if there were a Boolean "AND" between them like Google, but other databases have different rules, like treating multiple terms as a phrase search or as if there were an implied proximity operator (n4) between them. Web searches retrieve results by relevance, but Follett and WebFeat federated searches don't. Although results can be sorted by relevance in WebFeat and Gale PowerSearch Plus, the relevance ranking can be misleading for large results sets. The ranking algorithms only work on the first set of retrieved (often by chronological order) results, which may not be the most relevant articles available in a database. Queries including quotation marks for phrase searching, wild cards like "*," and proximity operators like "n4" are generally passed through by the federated engines but aren't supported in every database. Certain databases require "AND" as opposed to "and" in queries, and one database used "AND NOT" instead of "NOT." These types of inconsistencies between databases make simple Boolean queries the most consistent with results in the native databases in a federated search.

Federated search engines vary widely in the numbers of connectors to databases that are available for configuration (sixty-one to almost nine hundred in the tested engines). But having a connector on a list is not a guarantee that it will work. There are still many subscription databases that do not support federated searching, so it is unlikely that every source a school subscribes to will be searchable. The search time increases with the number of databases searched, but some products such as Gale's PowerSearch Plus, WebFeat, and OneSearch show interim results. Vendors claimed capabilities like OPAC searching before they were deliverable.

Federated search engines do not perform equally on the same databases. Follett One Search and WebFeat Express consistently performed more reliably (meaning retrieving the same results as the native interfaces) on most databases, but WebFeat Express picks up results from only one of six tabs in two of the test databases that have tabbed browsing, and six connectors would need to be configured for each database to retrieve content from all tabs. Follett One Search was the only search engine to retrieve all of the tabbed content. The Gale federated search performed the worst on its own (Gale) history databases, retrieving only a few, often-inexplicable results from almost every search. Gale also only retrieved the first set of results (fifteen) from World Book in every case, and is quite limited in the number of hits it will retrieve from any database. Follett One Search was not able to differentiate between ProQuest databases in order to search one individually.

The federated engine's querying abilities are different. While both Follett One Search and WebFeat Express allow for advanced searching features like subject or author searches and pull-down Boolean operators, and in most cases transmit advanced searches that replicate the same search in the native advanced interface, Gale PowerSearch Plus does not. Though it is possible to enter an advanced type of query in the basic or advanced PowerSearch interface, a basic keyword search string stripped of any of the Boolean operators will be transmitted to the federated databases when the federated search is

activated regardless of how the searcher entered the query. This serious flaw yields completely different search results than if the same Boolean query was entered as a search string in the federated interface.

System Configuration results

The process of configuration was straightforward for each product. The challenge in configuring the systems for this study was selecting a group of regularly used databases that all three products could search. Although all the major vendors are represented in each federated search product's list, there were many gaps and differences in the specific databases that are configurable for each vendor family. Neither of the two science databases (Gale and EBSCO) GCDS subscribes to is configurable for all of the products. There are still many databases that are not searchable by XML or Z39.50, the major federated search and retrieve protocol standards. Additionally, it was found that even if a connector configuration exists on a list, that doesn't mean it will work. A popular database for younger students called Sirs Discover was configurable in each of the search engines, but worked in none of them. Sirs technical support revealed that it is neither XML nor Z39.50 compliant, and can't be cross-searched, yet a connector was configurable in each federated product. WebFeat by far has the most comprehensive list of supported databases (almost nine hundred) because in addition to Z39.50 and XML, they can use an HTML screen scraping technique to search databases that are not Z39.50 or XML searchable. Gale PowerSearch Plus is the newest product, and has the fewest connectors available, 61 as of February 2008 but will probably continue to bring new connectors online. Follett One Search is the least expensive, but has been around for longer than Gale, and has about 175 available connectors.

An unexpected difficulty was searching the library OPAC. The OPAC search is integral to Follett, and was advertised by both Gale and WebFeat, but it was not ready to go for Gale PowerSearch Plus and WebFeat Express as advertised. WebFeat delivered a workable OPAC search after about a week, but Gale didn't get their advertised OPAC configuration to work at any point during the configuration and testing schedule.

WebFeat Express is significantly different from PowerSearch Plus and One Search in the amount of user customization of the interface that is possible. WebFeat Express can be configured with any of four different page layouts, and any number of preselected database groupings to help guide users to appropriate resources for either subject areas, grade levels, or even projects. "Branding" the search page with the banner from the school's website and customizing the colors to match was not difficult. It was also simple to generate the code to insert a customized simple search box (configured with preselected sources) into any webpage. Each feature was relatively easy to configure and worked right away. The only customization available for the other products is the selection of the one set of default databases to be searched.

Results of Systematic Search

The quantitative results for the eight series of queries in each federated search tool and each database native interface were recorded and analyzed. Observations from each series are numbered and reported below.

Search Series 1. Interpretation of the wild card "."* "Harlem rena*sance" and "harlem renaissance" results and findings: (a) The interpretation of the wild card is first a function of the database being queried. For example, Britannica and HNYT don't allow for a wildcard, so those queries retrieved no results. World Book skips the wildcard, and interprets the query as harlem renasance, a misspelled word. World book automatically executes a search using a word it determines as close, retrieving in this case, relevant results. (b) Gale PowerSearch Plus misinterprets the "*" query only in its search of History Resource Center both U.S. and World retrieving no results, but the query worked fine in its query to other databases. Executing the identical search in the native interface retrieved 108 U.S. and 30 World results, indicating that "*" is supported, so it is most likely a problem with the software connector to those databases. (c) The difference in retrieval numbers across the interface for the History Resource Center points to another significant difference in how results are treated by the three vendors. The History Resource Center native interface includes tabbed browsing (Reference, Biography, Periodical, News, Primary Resources, and Multimedia) of results. WebFeat only provided results from the default tab, in this case reference. Follett was the only engine that picked up all of the results from all of the tabs. (d) WebFeat appears to have some de-duping capability. (e) The search on harlem renaissance revealed a significant problem with Gale's federated searching of its own history databases. While Follett retrieved all the same results as the native interface and WebFeat retrieved all the results of the reference tab of the native interface, PowerSearch Plus retrieved only a small subset (4/105 U.S., 3/30 World) of the possible results. The PowerSearch results were inexplicable; they were neither particularly relevant, nor drawn from any one tab. (f) HNYT results shows that the Follett and WebFeat federated searches were retrieving more results than the same query in the native HNYT interface. The retrieved results showed hits from many newspapers, not only the HNYT. Two different problems were revealed. GCDS access to certain ProQuest databases comes from individual subscription and others come from the iConn consortium. They initially were set up with different authentication passwords. Fixing the authentication solved the problem with WebFeat (which had been searching other ProQuest databases—but not HNYT), but did not improve the Follett results. Follett's connection to ProQuest doesn't differentiate between databases, and will search all the ProQuest databases we can access—therefore retrieving essentially irrelevant results from the HNYT database. (g) With large result sets, it becomes apparent that PowerSearch Plus retrieves substantially fewer results. PowerSearch Plus will return up to three hundred results (the user can set the maximum to one hundred, two hundre, or three hundred) from the set of databases searched in its federated product. If there are five databases selected for searching, each can return at most the top sixty hits. Most of the searches were conducted with the results set limited to two hundred (or forty results per database). (h) HNYT results are returned in chronological order, meaning that the likelihood of finding a historical piece this way is small.

Search Series 2. Examining how search terms are executed. The focus of this study was to determine if the federated products were able to translate a Google-type keyword query into something usable by typical database syntax requirements. "Ancient egypt women," "Ancient egypt AND women," "Ancient egypt and women," and "'ancient egypt' and women" results and findings: This search revealed

tremendous differences about how a search is executed. (a) In a Google search the three terms are treated as though they have a logical "AND" between them. The results from the Britannica and World Book encyclopedias show a similar approach. The three keywords "ancient Egypt women" and the search "ancient Egypt AND women" return identical results, but only if "AND" is used and not "and" (for Britannica). However, "ancient Egypt women" retrieved zero results from the Gale and Ebsco history databases in either the native or federated interfaces, indicating that there is no translation of Google-like keyword searching (implied "AND" between words) into database search syntax ("AND" required between terms) occurring in the federated search engine. (b) Encyclopedia Britannica will not allow quotation marks for phrase searching. The only way to conduct a proper phrase search is through the advanced search interface in Britannica. This produced thirty-eight excellent hits in the native interface, but couldn't be replicated in either a basic search-box search or in the advanced (three text boxes with pull down Boolean operators between) options available with any of the federated engines. (c) PowerSearch Plus will automatically modify queries. If you enter the query "ancient egypt and women" in the regular PowerSearch interface and click on the additional databases (federated) tab to activate the federated search feature, the query changes to "ancient egypt women" in the search box. This was detrimental to this query's effectiveness.

Search Series 3. Exploring Boolean "NOT." "Red scare NOT mccarthy" results and findings: (a.) The HNYT database uses the more unusual syntax "AND NOT" for a Boolean "NOT" search, resulting in zero hits. Conversly, executing an "AND NOT" search in all the databases resulted in zero hits in all but HNYT and Britannica. It appeared that Britannica interpreted "AND" and "AND NOT" the same way. (b) Again, Gale's federated search of its History Resource Center (U.S. and World) databases showed some inexplicable results. The discrepancy in the number of hits could not be explained by either numerical limits or by retrieval of only one tab of results. A number of searches conducted in the native interfaces failed to replicate the PowerSearch Plus results. The small set of results included articles from multiple tabs, but not the best articles available. The results were repeatable. (c) PowerSearch Plus showed some evidence of translating the query to World book. The three results available for that search were replicated in the native World book interface by using the advanced search option.

Search Series 4. Examining how searches might be executed differently through one single search box versus advanced interface with three text boxes plus pull down Boolean operators and Boolean "OR." "Desegregation AND busing," "desegregation OR integration AND busing" as single search string and seperated in advanced search results and findings: (a) The "OR" operator expanded the search as expected in all cases. (b) PowerSearch Plus federated search offers no advanced search interface, but it is possible to enter an advanced search query in the regular PowerSearch interface, which transfers a basic search string to the PowerSearch Plus interface (minus the Boolean operators!) The automatic removal of the Boolean operators from a query entered in the regular PowerSearch interface when it is transferred to the federated engine is a serious problem. (c) Follett One Search and WebFeat offer advanced search interfaces. WebFeat's queries were consistent with the same queries in the databases' native advanced search interfaces (bearing in mind it only picks up hits from one tab in the History Resource Centers U.S. & World). It returned the same results as the keyword query executed the same way in the native interface regardless of whether it was executed in the single search box or in the advanced interface. Follett One Search returned the same results for the queries executed as a single search string or as an advanced search, but for Gale history products the results closely matched the search executed as a single search string in the native interface, but not the advanced search string. (d) Again, PowerSearch Plus demonstrated some inexplicable results when searching the Gale history databases; its results for History Reference Center and HNYT reflect the truncation at 200/5 or 40 results, which was expected. (e) Follett One Search's connector failed to translate its advanced query into a valid query for Britannica.

Search series 5. Misspelled query. "Harlem renasance" results and findings: (a) In both the native interface and a federated search, World Book automatically replaces the misspelled word with a different word, and in this case produced relevant hits. No other databases returned results. (b) Upon entering a misspelled query in the basic PowerSearch interface, several "did you mean" alternatives are offered, but if you are already in the federated tab and a misspelled query is entered, the misspelled query is broadcast as is with no corrections offered. (c) Follett offers a "did you mean" function only for its library search; it doesn't work for a federated OneSearch. (d) WebFeat offers a "did you mean" service upon entering a misspelled query. (e) Native database interfaces offered some sort of "did you mean function," or related subject heading choices. The Encyclopedias both didn't; Encyclopedia Britannica offers a choice to automatically expand the query, which leads to thousands of irrelevant hits, and World Book provided automatic correction.

Search series 6. Proximity operators ("reform n5 tax," "reform tax," and "tax reform"). An n5 proximity operator should find any hits that have both words, within five words of each other in any order. One would expect the query "reform n5 tax" (or even "n2") to find all of the hits of the more frequently used "tax reform" plus at least some additional hits. Results and findings: (a) Only three of the possible seven data resources allow proximity operators as a part of their supported search syntax, yielding no hits for either encyclopedia, the HNYT, or the OPAC. (b) Follett One Search and WebFeat Express both translated the query exactly as it was entered, and the results from the EBSCO History Reference Center, which supports proximity searching, returned the same results as the native interface across the board (accounting for Gale PowerSearch Plus returning at most thirty-nine or forty hits) with the "n5" search returning more results (192 versus162) than the search for "tax reform," which was expected. (c) The two Gale history databases showed more results in the Follett One Search than in the native interface. The reason for this was determined to be that History Reference Center native interface search retrieves at most two hundred results in its default "relevance search." However, Follett One Search federated retrieves results from this database in chronological order, which is not limited to two hundred hits and results in higher numbers of results than in the native interface. The low number of WebFeat Express hits again are due to its retrieval of only one tab's results. (d) Comparing the results of the two Gale history databases in this series shows that a two-word query is searched as an implied proximity search. The implied n4 default was confirmed on the search tip page in the native database. The "n5" search picked up a couple additional hits In Follett One Search and WebFeat Express as expected. (e) Gale PowerSearch Plus again showed inexplicable retrieval of results from the Gale history databases. (f) HNYT treats two-word queries as a phrase search by default.

Search series 7. Subject searching. The strings "world war ii 1939-1945," "world war 1939-1945," "world war 1939 1945, and "world war 1939 1945 AND causes" were executed as an advanced subject and keyword search combination, and as a basic keyword search string. This series explores the use of subjects as either keyword or subject searches. Results and findings: (a) Different databases use different subject headings: Gale uses "world war ii 1939-1945" as a subject heading, EBSCO and Follett Destiny use "world war 1939-1945." The first search failed only in the EBSCO History database. (b) Gale

PowerSearch Plus retreived zero results for either search in either of the Gale history databases. It was determined the search failed in this case because of the dash in the date 1939-1945. The query "world war 1939 1945" retrieved some results. Interestingly, if the query was entered into the original PowerSearch interface, the dash would have been eliminated as the query was transferred to the federated interface. (c) Using the following queries for "subject=" or "SU:" retrieved no results in any cases. (d) Using the advanced subject search pull-down was equivalent to a keyword string search in Webfeat Express only. (e) A subject search entered into Gale PowerSearch Plus advanced interface doesn't get translated properly to its targeted databases, a keyword search performs better.

Search series 8. Author search vs. keyword search. "Thomas Friedman" was executed as a keyword and as an author search from the advanced search pull-down. Results and findings: (a) The data show that Gale PowerSearch Plus transfers a keyword search query to the federated databases regardless of whether the searcher uses the author search or a keyword search in their original query. Follett One Search and WebFeat Express appear to transfer author queries as author queries and keyword queries as keyword queries.

Discussion

Well designed and thoughtfully configured federated search tools can provide a simple interface to access multiple quality database sources and the OPAC in a single search for students who either lack, or don't bother to use, sophisticated information literacy skills. Searchers will get the best results with simple Boolean queries if they include the operator "AND' in their multi-word queries, or use the advanced interface to phrase their queries, requiring a certain level of information skills instruction. Federated search tools will continue to be imperfect and slower than Google as long as the architecture remains based on distributed live searching, but the ease of obtaining quality content they provide might entice student users away from Google, particularly if their teachers require authoritative sources.

Federated Searching in the School Library

Ideally, there would be a place for comprehensive information literacy curriculums in all primary and secondary schools. In such a forum, students could be exposed to the breadth and depth of library resources and learn how to skillfully search all types of information resources. Unfortunately, not enough students are being exposed to such curriculums, even where they exist, for a variety of reasons. Flexible scheduling and collaborative research projects work well, but not all grade levels and teachers will work with the library, leaving large gaps in the number of students reached. Schools also differ in resources and professional staffing. Since federated searching best serves novice searchers with little exposure to information literacy training, it may have a place in many schools.

Federated Search Engine Performance

The systematic searching study revealed important differences in how federated searching differs from searching the Web and from searching in native database interfaces, but doesn't tell the whole story of these products because they are so fundamentally different from each other. Below, each product's pricing and general performance is discussed. The appendix tabulates the features and functionality of the three federated search engines for a more in-depth overview of the range of the federated search engine's capabilities. It must be understood that the findings in this study represent hosted federated search product capabilities at the time of testing, and that products will continue to add new and fix broken connectors, and add functionality and improvements as they mature.

Gale's PowerSearch Plus did not perform well in this study because of its limited number of retrieved results, inexplicable results at times, its inability to deliver an OPAC search, and because of its poor performance on the Gale History Center databases. Most of these issues are related to Gale's emphasis on clustering and ranking retrieved results, which was not evaluated as a part of this study. To cluster and rank, all the results must first be retrieved by the search engine, then be subjected to a ranking algorithm. This process is very time-consuming in real time, and as a consequence, can only reasonably work on limited (up to three hundred) numbers of results. To keep the process moving along, queries to databases can time out, meaning that according to Gale technical services (Michaela 2008), a database will return fewer or no results if it times out before the search is complete. The response time for the Gale History Resource Center databases may have been responsible at least in part for poor retrieval numbers.

Gale's federated functionality is not well integrated into the PowerSearch platform. One must first execute a search in the basic PowerSearch interface (which is completely different), and then click on the additional databases tab in the results section to activate the federated search engine. This could be a business decision to ensure that Gale resources will be seen first by the searcher since results from other vendor's products require additional steps to retrieve. A serious design flaw was detected in this approach. Any Boolean operators used in the original query entered in the basic PowerSearch interface are stripped from the federated query, yielding a completely different query than originally intended, and an inferior set of results for the test case. Interim results are not shown, only an indicator of progress, and it seems slow, especially on wireless. The results, a clustered outline and visual map, are really great features, but at $1,000 per year, this new product clearly has some kinks that need to be worked out.

Follett One Search, at $499 per year, is the least expensive integrated search tool tested. It offers a reliable and comprehensive search, but offers no opportunity to post-handle results. It is integrated seamlessly into the Follett Destiny search environment and works well with either a basic or advanced search. It is easily configured, and can also search many great free resources. The librarian can select a default set of sources to be searched, but the user can change the list. The connector list is not an exhaustive one. A minor problem with OneSearch was that while it does show interim results, very often the

interface would fail to show that a search was complete (it would appear to be still searching even after several minutes even though the search was actually completed after 10 to 20 seconds). The user would have to click "get results" to get the rest of the results.

At around $8000 per year, WebFeat Express is the most expensive federated search engine tested for the school library, yet it is the most capable of providing an effective *single* interface for all of a library's searching needs including the OPAC. It offers the most connectors, ample customization features that are reasonably configurable by a librarian, and it has the most capability for manipulation and refining of results. Because of its capability to provide multiple predefined sets of resources and predefined simple search boxes, it is by far the most adaptable. One drawback is that for databases with tabbed results like the Gale History databases, a connector for each tab must be separately configured. To completely configure both history databases would use twelve connectors out of the fifty included in the base price. Tabbed browsing is not an uncommon database feature.

Future of Federated Searching

Federated searching will not be more Google-like until the current live database search model is abandoned in favor of a centralized search against pre-built indexes similar to Web search engines. Although this is technically and economically feasible, the database publishers must have a reason to allow access. This is already occurring for scholarly content as Google Scholar paves the way toward the future of federated searching in the academic library by its continuing pursuit of access to crawl and index proprietary databases of scholarly content. Many nonscholarly database vendors (the school market) are lagging in both their support of federated search and retrieval standards Z39.50 and XML, and allowing access for indexing. As a result, a school must be careful in selecting a federated product that can support a search of their most heavily used database products.

Areas of Future Research

Federated searching by school-aged children is an area in which little research has been done. Studies focusing on the implementation of federated search engines in schools and their application to the information requirements and searching abilities of students are a needed next step. Other areas of interest include best practices and instructional activities for teaching federated searching, and the effect of federated search engines on students' overall information literacy skills. As federated searching products continue to mature, and more information providers support federated search and retrieval standards, further systematic studies of federated searching products in schools should continue.

Works Cited

Breeding, M. 2005. Plotting a new course for metasearch. *Computers in Libraries* 25(2): 27–29. ERIC database (accessed Feb. 15, 2008).

Curtis, A. M., and D. G. Dorner. 2005. Why federated search? *Knowledge Quest* 33(3): 35–37. Academic Search Premier (accessed Feb. 21, 2008).

De Rosa, C., J. Cantrell, J. Hawk, and A. Wilson.2006. *College students' perceptions of libraries and information services*. www.oclc.org/reports/pdfs/studentperceptions.pdf (accessed Feb. 18, 2008).

Fahey, S. 2007. F******ed searchers? The debate about federated search engines. *Feliciter* 53(2): 62–63. EBSCO Professional Development Collection (accessed Feb. 8, 2008).

Freund, L., J. R. Nemmers, and M. Ochoa. 2007. Metasearching: An annotated bibliography. *Internet Reference Services Quarterly* 12(3/4): 411–30. Haworth Information Press database (accessed Feb. 14, 2008).

Frost, W. J. 2004. Do we want or need metasearching? *Library Journal* 129(6): 68. Academic Search Premier (accessed Feb. 14, 2008).

Labelle, P. R. 2007. Initiating the learning process: A model for federated searching and information literacy. *Internet Reference Services* 12(3/4), 237–52. Haworth Press database (accessed Feb. 14, 2008).

Large, A. 2004. Information seeking on the web by elementary school students. *Youth information-seeking behavior.* Chelton, M. K. and C. Cool (Eds.): 293–319. Lanham, Md.: Scarecrow.

Levin, D., and S. Arafah. 2002. *The digital disconnect: The widening gap between internet-savvy students and their schools*. http://epsl.asu.edu/epru/articles/EPRU-0208-36-OWI.pdf (accessed Feb. 9, 2008).

Minkel, W. 2002. Metasearching comes of age. *School Library Journal* 48(1): 33–34. Academic Search Premier (accessed Feb. 15, 2008).

NISO Standards Committee BC Task Force 3. 2006. *Metasearch XML gateway implementers guide* (NISO RP-2006-02). http://www.niso.org/committees/MS_initiative.html (accessed Feb. 15, 2008).

Sadeh, T. 2007. Transforming the metasearch concept into a friendly user experience. *Internet Reference Services Quarterly* 12(1/2): 1–25. Haworth Information Press database (accessed Feb. 14, 2008).

Solomon, M. (2004). A confederacy of databases. *Searcher* 12(7): 25–29. Academic Search Premier database (accessed Feb. 20, 2008).

Tenopir, C., B. Hitchcock, and A. Pillow. 2003. *Use and users of electronic library resources: An overview and analysis of recent research studies* (CLIR Rep. No. 120). www.clir.org/pubs/abstract/pub120abst.html (accessed Feb. 9, 2008).

WebFeat. n.d. History. *Webfeat: The Original Federated Search Engine*. www.webfeat.org/company/history.htm (accessed Feb. 26, 2008).

Young, T. E. 2005. Federated searching: Welcome to the era of the super search. *School Library Journal* 51(12): 95. Academic Search Premier (accessed Feb. 15, 2008).

Appendix. Analysis of Federated Search Product Features

Feature	One Search	PowerSearch Plus	WebFeat Express
Resource Discovery			
Number of clicks to federated results	1 or 2 (sometimes necessary to click “get results” to return all federated results)	(1) Submit query in standard Power Search interface (2) Click in "additional databases" tab to start federated (Plus) function.	(1) User must select at least one resource group to search (2) submit query
User selection of resources to search? Limit?	Yes, click to select or deselect listed configured databases.	Yes, but click to see selected databases. Also, one searches all available Gale DB's or none, no choices among them.	Yes, Subject groups are configurable, which creates tailored groups of databases for different subjects, grade levels, etc.
Configurable default set?	Yes, default set	Yes, default set	Yes, any number of sets

Feature	One Search	PowerSearch Plus	WebFeat Express
Resource Discovery			
Single search box available to place on any webpage?	No	Yes, but for the basic PowerSearch interface, not the federated part	Yes, any number of differently preconfigured search box codes may be generated and used in webpages.
Limit of number of databases configured			50
Limit number of databases searched			50
OPAC Search?	Yes, Integral	No, (not yet?)	Yes

Feature	One Search	PowerSearch Plus	WebFeat Express
Query			
Pull down Boolean operator "advanced" search available?	Yes	Not in the federated function, regular PowerSearch only	Yes
Single search box?	Yes	Yes	Yes, basic, but with pull down choice of keyword (def.), title author, subject, all
Any limiters available?	No	Refining after initial search results	Publication date (year) range, full text

Home authentication?	Yes	Yes, once	Yes, once
Browsing			
Initial results shown?	Yes	No	Yes
Indicator of progress?	Yes, but doesn't always indicate that the search is complete	Yes	Yes

Feature	One Search	PowerSearch Plus	WebFeat Express
Browsing Results			
Errors displayed?	Follett errors are displayed	No	No
Can results be refined? How? Type of doc, source, pub., keyword, date etc.	No refining—results are presented as they are retrieved	By source, by keyword, by date	No refining
Can results be ordered/sorted? How? Is there relevancy ranking?	No manipulation of results	Results are clustered in a visual map or outline presentation view. Relevancy ranking based source rankings	Grouped by source by default or sorted by title author date, cluster, or relevancy
Characterize brief description of item	Almost complete citation: graphic previews, title, author, journal, volume, issue, page, issn, full text PDF or HTML, reading level. Description for encyclopedia entries.	Default view shows title, URL, and source, more detail shows author and rank in addtition, less detail show only title.	Title, author, journal, pub date, database source, sometimes description of work, graphic previews if available, or subj. headings depending on source.

Feature	One Search	PowerSearch Plus	WebFeat Express
Browsing Results			
Is more info available? from target source? Or within metadata?	Link to actual item	Selectable detail "more, less, or normal," & link to actual item	Link to actual item
Any clustering or catagorizing of results?	No—by source only	Clustering (visual or outline) & relevance by Grokker	Yes, clustering and relevance
Indication of full text?	Yes	No, but didn't return non–full-text articles	Yes
Can results be ordered/sorted? How? Is there relevancy ranking?	No manipulation of results	Results are clustered in a visual map or outline presentation view. Relevancy ranking based source rankings	Grouped by source by default or sorted by title author date, cluster, or relevancy

Feature	One Search	PowerSearch Plus	WebFeat Express
Retrieval			
Any partial retrievals? Items found vs.	Difficult to verify. One sometimes has to	If a query to a database is not retrieved	No

items retrieved? Are they updated?	click the get results button to get results, but ostensibly, it may still have been gathering results.	quickly enough, it will time out retrieving fewer or no results.	
How are very large sets of hits handled?	Lists number of results available	Retrieves total of 100, 200, or 300 results (200 default) from all databases queried	Shows total number available, immediately retrieves only first set.
number of items retrieved limit?	No	If total is 200 and 5 sources are selected each can retrieve a max of 40	No
E-mail results	No	Yes link	Yes, citation info, but no link or even database source description

Feature	One Search	PowerSearch Plus	WebFeat Express
Retrieval			
Save/export results?	Can save to a list if user is logged into destiny as a user.	Book mark, or post to del.icio.us	Yes, as RIS file or direct export to RIS Format file (ProCite, EndNote, Reference Manager
Generate citations	Citations can be generated in destiny from lists. No specific format.	No	No
Services to other systems?	No	Post to del.icio.us	Direct export to RIS Format file (ProCite, EndNote, Reference Manager
User workspace?	If user logs in and creates lists	No	No